The Eyes Have It

Terry Riopka
Computer Science and Engineering
Lehigh University
Bethlehem, PA 18015
riopka@kantbelievemyeyes.com

Terrance Boult
Computer Science Dept.
Univ. of Colorado at Colorado Springs
Colorado Springs, CO 80933
tboult@cs.uccs.edu

ABSTRACT
This paper evaluates the impact of eye localization on face recognition accuracy. To investigate its importance, we present an eye perturbation sensitivity analysis, as well as empirical evidence that reinforces the notion that eye localization plays a key role in the accuracy of face recognition systems. In particular, correct measurement of eye separation is shown to be more important than correct eye location, highlighting the critical role of eye separation in the scaling and normalization of face images. Results suggest that significant gains in recognition accuracy may be achieved by focussing more effort on the eye localization stage of the face recognition process.

Categories and Subject Descriptors
I.4.m [**Image Processing and Computer Vision**]: Miscellaneous

General Terms
Measurement, Performance, Experimentation.

Keywords
Face recognition, eye localization, biometrics, PCA, EBGM, FaceIt

1. INTRODUCTION
Face-based biometrics have been growing in importance and researchers continue to seek more advanced algorithms to improve face system performance, especially under difficult conditions. Most face recognition systems rely on accurate detection of facial features either as input to the classifier directly, or more commonly for the purpose of normalizing images [4]. This paper suggests that when looking for ways to improve face biometrics, the eyes have it: eye localization is currently a major weak link.

Symmetry of the eyes and their consistent relationship with respect to other fiducial marks on faces make them extremely useful for parameterizing and normalizing geometric features of the face. Because eye separation does not change significantly with facial expression, nor with up and down movements of the face, eye separation distance is often used for face normalization. Nose distance, another feature often extracted, is relatively constant with respect to side to side movements of the face and also depends on accurate eye localization. In addition, orientation of the line between the eyes is often used to correct for pose variations. Lastly, eyes are essentially unaffected by other facial features like beards and mustaches, making them invaluable features to most face recognition systems. It is therefore not surprising that up to 1/3 of the total processing time for many face recognition algorithms is consumed by the eye localization stage [1] further underscoring its importance.

In appreciation of the critical importance of eye localization, this paper investigates the question: **what effect does the accuracy of eye localization have on face recognition accuracy?** Using several face recognition algorithms, we present an eye perturbation sensitivity analysis as well as empirical evidence using real images under various weather conditions that show that eye localization has a significant effect. Clearly, this effect will vary somewhat from algorithm to algorithm but our contention is that all face algorithms will be comparably affected given the vital role that eyes play in face recognition. We suggest that efforts made to improve recognition accuracy may be better spent improving the eye localization stage rather than improving the classification algorithm itself.

The paper is organized as follows. Brief descriptions of the face recognition algorithms used is presented, followed by a sensitivity analysis with respect to a systematic perturbation of eye coordinates. Next, we describe the set-up used to acquire face images under varying weather conditions and present results showing the effect of eye localization on algorithm performance. Results are also presented for face images under varying illumination and pose. We conclude with a discussion of the ramifications of our study and suggest avenues for further research.

2. FACE RECOGNITION ALGORITHMS
Three different face recognition algorithms were investigated in the following experiments: Principal Components Analysis (PCA) [17], Elastic Bunch Graph Matching (EBGM)[10] and FaceIt, a commercial application based on an LFA algorithm [11]. PCA and EBGM algorithms were provided by the Colorado State University (CSU) Face Identification Evaluation System (Version

WBMA '03, November 8, 2003, Berkeley, California, USA.

5.0) [2]. FaceIt was implemented using programs built from a software development kit licensed from Identix Inc. Brief descriptions of each algorithm are provided below, but the reader is referred to the relevant publications for details.

Principal Components Analysis (PCA)
The PCA recognition algorithm is a nearest neighbor classifier that operates in a subspace whose basis vectors are the eigenvectors of a scatter matrix formed by training images. Feature vectors are formed from training face images by concatenating the image pixel values and subtracting the mean vector. This process generates a set of very large, highly correlated vectors which are then rotated into a much smaller subspace with no sample covariance between features. During training, the most significant eigenvectors and the mean training vector are stored. A novel image is recognized by subtracting the mean vector, projecting the result into the PCA subspace and then identifying the training image whose projection is closest to the input. The distance metric used in this paper was the Mahalonobis Cosine [2].

Elastic Bunch Graph Matching (EBGM)
The EBGM recognition algorithm [10] uses a wavelet transform to generate feature vectors at various landmark points in a face image. These feature vectors, known as "Gabor jets", are then associated with the nodes of a face graph created for each training image. Analogous to the projected vectors of PCA, the face graph serves to represent the face image in a low dimensional space. A novel image is recognized by creating a similar face graph for it, and, through a complex graph matching algorithm, measuring its similarity to stored face graphs in the training database.

FaceIt
While it is derived from LFA [11], FaceIt is a commercial product with years of unpublished improvements and modifications of improvements. The original LFA approach computes its features starting from a PCA space and transforms the resulting data using filters to produce more localized features. However, since some extracted features have an extent that span more than half the face, the features are not all that local.. The basic LFA features are dense in the original image space, and a more sparse subset is computed using a neural net. This sparse set provides for effective reconstruction, even for frontal face images not in the original training set, using a relative small number of coefficients. Because of the localized nature of the features, the approach is inherently more robust with respect to lighting variations. However, it is interesting to note that the Identix literature claims the system uses the relative distances between different landmarks on the face to create a facial biometric template which is used for matching in recognition.

3. EYE PERTURBATION EXPERIMENTS

Face recognition algorithm papers tend to focus on the recognition and representation components, but all of them, including the LFA and PCA approaches, depend on preprocessing that begins with localization of the face and the eyes. Using the eyes, and possibly other features, the image is normalized to a constant size and shape prior to feature extraction. The implication of error in eye localization is thus not just a shift of features, but also a scaling and possible rotation of the input image.

To investigate the effect of eye localization, we begin with a sensitivity analysis using perturbed eye locations as input to the three face recognition algorithms: PCA, EBGM and FaceIt. We take a non-standard approach and use the same set of images as both probe and gallery. This allows the face recognition system to match the actual input image. By allowing exactly the same image as both the probe and the gallery, we isolate eye localization - there is no variation in pose or lighting. While unrealistic in the sense that a face recognition algorithm would hopefully produce identical eye locations given identical images, the removal of all other factors allows us to focus on the impact of eye location alone.

3.1 Experiment Conditions and Factors

In this experiment, a subset of the full FERET database [13] was used to create a gallery consisting 1024 images, where we use 4 different front-facing images of each of 256 different subjects. Subjects varied in age, race, gender, with and without glasses. The same images were used as probes, but with perturbed eye locations input to the face algorithms. The perturbed eye locations were offsets relative to the known center of the eye for each gallery image as depicted in the 13x13 pixel grid shown in figure 1. For example, the top left black pixel represents the location of one eye offset six pixels to the left and six pixels up from its true location. For each algorithm, a total of 289 experiments were run, using a total of 17 perturbations per eye.

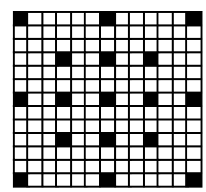


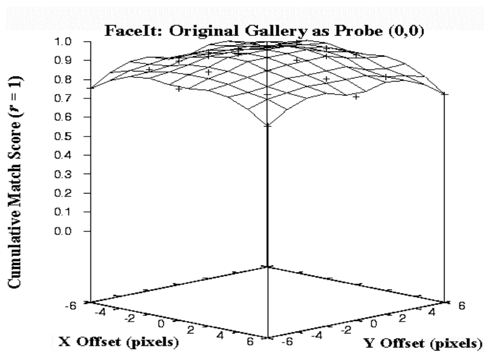
Figure 1. Eye perturbations depicted pictorially as black pixels in the 13x13 grid centered on the known location of the eye.

3.2 Performance Metric

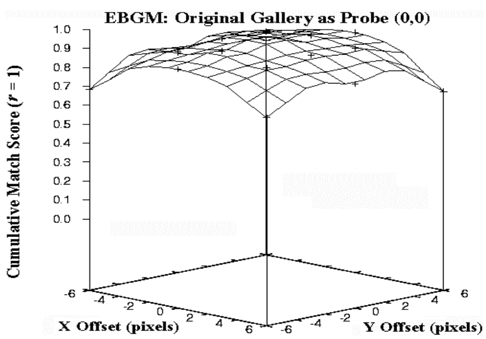
Performance was measured using a cumulative match score or CMS [13]. The CMS is a function of an independent variable r (rank) and a rank set E, and is defined as the fraction of E that has a rank of r or lower. It can also be viewed as the fraction of probes yielding a correct match within the top r candidates. The CMS was computed using the best rank obtained for the individual.

3.3 Results and Discussion

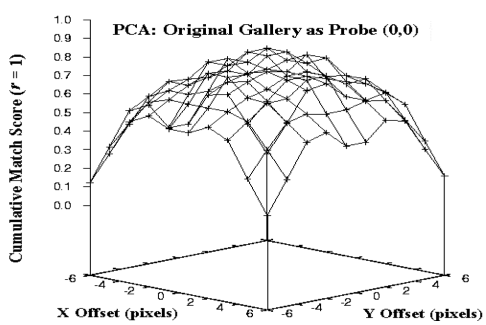
In all, 17 surface plots for each fixed eye could be shown, but we present two examples for each algorithm which show the range of possibilities. 3D mesh plots are interpolated to aid in visualization. The first set of plots has the left eye fixed at the correct location with the right eye systematically offset. The result, shown in figure 2, is that there is only a minor impact on the overall stability of the match for both FaceIt and EBGM. PCA is clearly more significantly affected by perturbations in eye loca-



(a)

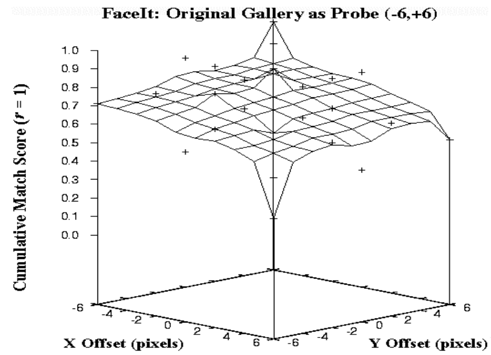


(b)

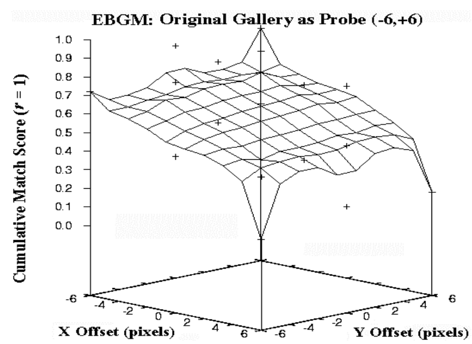


(c)

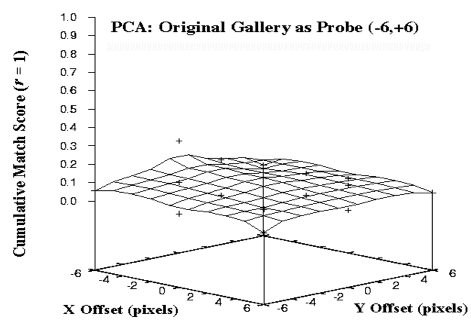
Figure 2: Performance (CMS) for FaceIt (a), EBGM (b) and PCA (c) as a function of right-eye perturbation, with left eye fixed at the correct location [0,0].



(a)



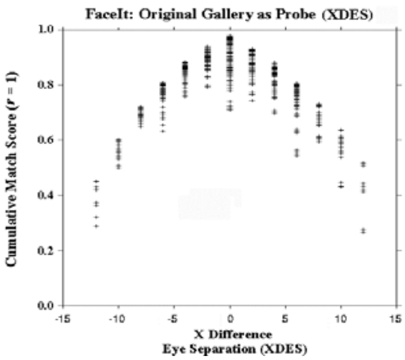
(b)



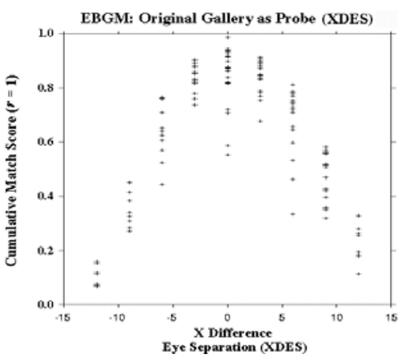
(c)

Figure 3: Performance (CMS) for (a) FaceIt, (b) EBGM and (c) PCA as a function of right-eye perturbation, with left eye fixed at [-6,+6].

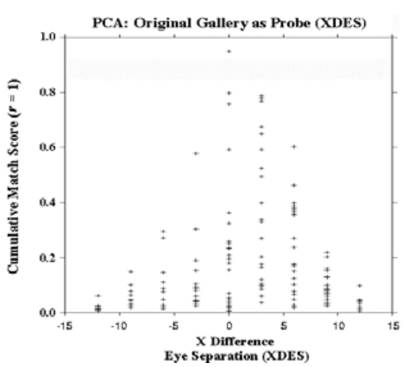
tion, no doubt a result of its greater dependence on face image alignment. This is evident with even greater clarity in the second set of plots of figure 3, showing results for the left eye displaced by 6 pixels in Y and –6 pixels in X. Since the left eye is displaced -6 pixels, figure 3 shows instances for which the Y offset varies and the total eye separation in x increases from 0 to 12 pixels. As expected, there is a significant reduction in the overall performance as the total displacement between the two eyes increases, with FaceIt appearing to be slightly more robust in this respect.



(a)

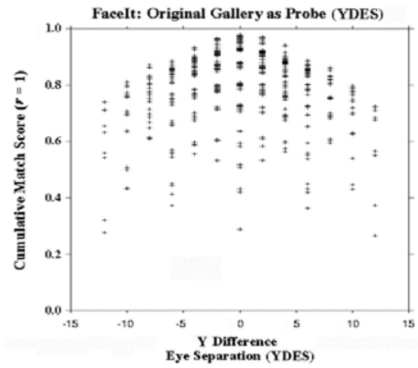


(b)

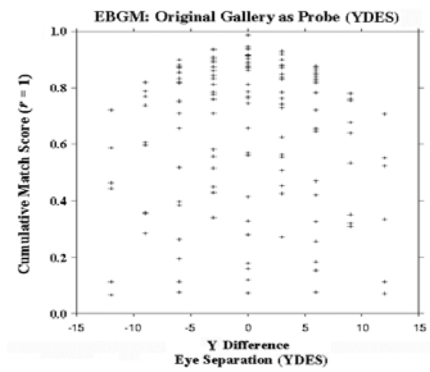


(c)

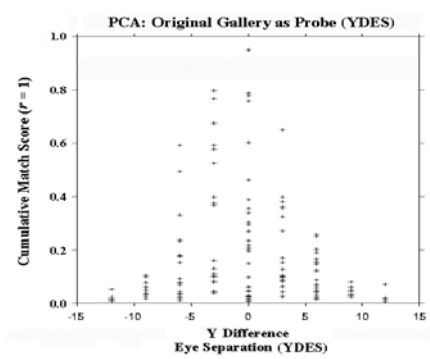
Figure 4: Performance (CMS) for (a) FaceIt, (b) EBGM and (c) PCA as a function of eye separation in the x direction (XDES).



(a)



(b)



(c)

Figure 5: Performance (CMS) for (a) FaceIt, (b) EBGM and (c) PCA as a function of eye separation in the y direction (YDES).

To better summarize the overall two-dimensional effect of offset,performance was also plotted against two computed variables: the difference in eye separation in the x direction (XDES) and the difference in eye separation in the y direction (YDES) (figures 4 and 5). Note, XDES is rather intuitive, negative for eyes closer together and positive for eyes further apart. Subtraction of y offsets (YDES) yields a value directly proportional to the slope of the line connecting the eyes, negative for the left eye above the right eye, and positive for the left eye below the right eye.

From figure 4 it is evident that FaceIt and EBGM performanceis almost equally degraded as the eye separation decreases or increases and is relatively independent of YDES. This is further supported by noting that even when the line connecting the two eyes is not horizontal, performance still depends mostly on XDES. This is seen in figure 5, where YDES is plotted as a function of performance. Note that even for absolute values of YDES that are high, when XDES is low (that is, eye separation remains about the same), performance can remain relatively high provided the difference in the separation of the eyes in the x direction is minimal. The effect of eye localization appears to be more dramatic when errors are made in eye separation distance versus errors made in identifying the actual locations of eyes.

Apparently, in instances where eye separation remains unchanged, but is shifted, performance is less affected in both FaceIt and EBGM algorithms. This is clearly not the case for PCA, which appears to be affected by both, and is severely degraded when the location of one or both eyes is in error (see figure 3c). In fact, this supports our contention that accurate eye localization with respect to normalization and scaling is vital to effective face recognition. PCA is more drastically affected due to its more global approach and weaker dependence on local features, resulting in a very strong dependence on face registration and hence eye localization.

4. ATMOSPHERIC IMAGE DEGRADATION

The following section describes a set of experiments performed to investigate the effect of eye localization under more realistic conditions, more specifically, for images degraded by atmospheric imaging effects. First, the perturbation experiments of section 3 are repeated using a single set of atmospherically degraded face images showing that "real" images yield similar behavior with respect to eye perturbations. Second, performance of FaceIt on a large number of atmospherically degraded images using FaceIt eye localization is compared to its performance using known eye coordinates, showing that accurate eye localization can make a significant difference in recognition rate.

4.1 Experiment Setup

To study the effects of weather on face recognition accuracy for our research, we have set up an outdoor LCD monitor on which we can project and acquire images of, face images (see figure 6). Images are projected on a 15" LCD marine monitor (300 nit display) and acquired asynchronously by two cameras at high speed from a distance of 94 and 182 ft. A single gallery of images (as described in section 3.1) is projected and acquired every day, approximately 15 times a day. All images are marked with a pair of easily identifiable markers so that regardless of the capabilities of the recognition algorithm, correct eye locations can be computed. Given the eye locations of the original FERET images and being able to locate the image markers enables us to accurately compute the location of eyes in all of the projected images. This apparatus was used to acquire experiment data.

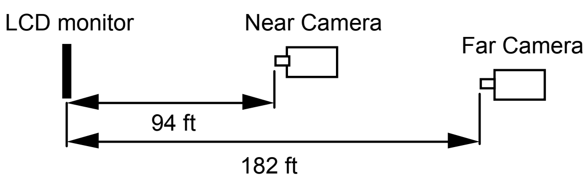


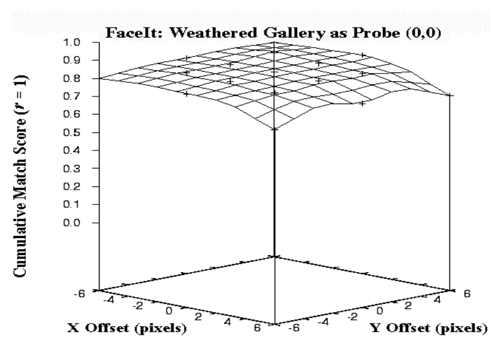
Figure 6: Camera setup for atmospheric degradation experiments.

4.2 Eye Perturbation Experiments Using "Weathered" Images

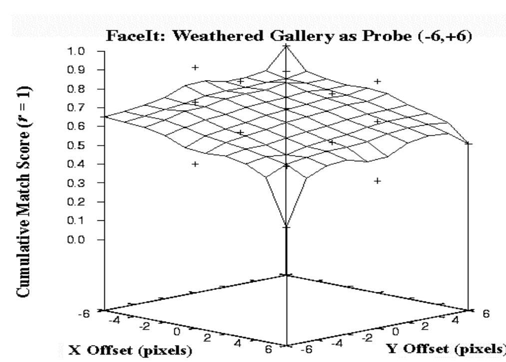
The goal of this experiment was to examine the impact of eye-localization when there is actual imaging system degradation in the probe images.

Experiment Conditions and Factors. The same set of images as used in the perturbation experiments was used here except that probe images consisted of the entire image set projected outside and re-acquired. The particular image set used was selected somewhat arbitrarily from a day with clear weather (*i.e.* no precipitation, snow, fog or mist). Experiment runs over the same set of eye perturbations were made for all algorithms.

Results and Discussion. Results are shown only for FaceIt, since similar observations were made for all three algorithms. Note the similarity between figures 7a, 7b and 2a, 3a respectively. At a crude level, atmospheric degradation is analogous to a mixture of blurring and noise being added to the experiment. Thus it is not surprising that the behavior with respect to eye perturbations appears to be similar, but smoothed out. In other words, while blurring tends to stabilize the algorithms with respect to eye localization errors, their general behavior with respect to a strong dependence on eye localization remains unchanged. This is evident even in the x and y difference plots in figures 8a and 8b (compare to figures 4a and 5a). Similar analogous results were observed for both EBGM and PCA (not shown).

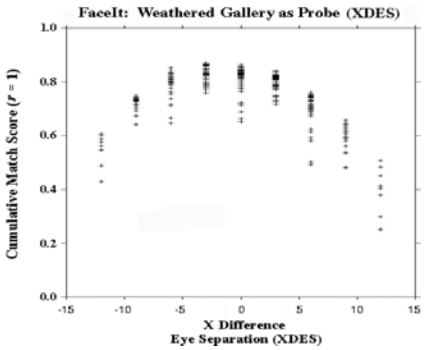


(a)

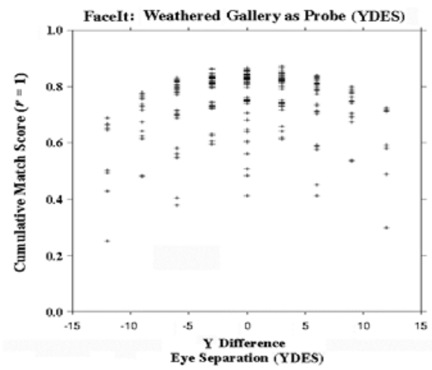


(b)

Figure 7: Performance (CMS) for FaceIt on "weathered" face images as a function of right-eye perturbation, with left eye fixed at (a) [0,0] and at (b) [-6,+6].



(a)



(b)

Figure 8: Performance (CMS) for FaceIt on "weathered" face images as a function of eye separation in the (a) x direction (XDES) and (b) y direction (YDES).

4.3 FaceIt Experiments Using "Weathered" Images

The goal of this experiment was to examine the effect of eye-localization error on the performance of a leading face recognition algorithm on real images acquired in the field. FaceIt was chosen due to its standing as one of three best performing face recognition algorithms in FRVT2000 and FVRT2002 [12]. It was also fast enough for the large number of face recognition experiments required by our experiments.

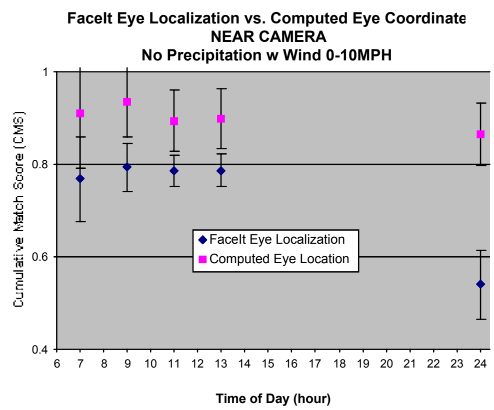
Experiment Conditions and Factors. Again, the same set of images as used in the perturbation experiments was used here except that in the analysis, the gallery consisted of every first image in the set (known from previous experiments to be the "best" out of the four). Each probe set (one for each camera) consisted of images taken at various times of day, during days with no precipitation, snow, fog or mist. All images were acquired over approximately three months in the spring of this year.

Performance Metrics. As before, the cumulative match score (CMS) at rank 1 was used as a measure of performance. However, in this case, four CMS values per experiment were averaged to reduce the overall variance.

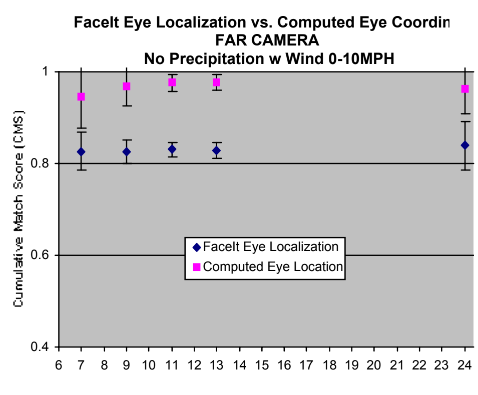
Results and Discussion. Results are shown in figures 9a and 9b. CMS measures are shown for 5 times of day for both near and far cameras. Images between 2:00pm and 6:00pm were omitted due to abnormal imaging effects caused by angle of the afternoon sun. This resulted in 59, 63, 56, 47 and 101 probe set samples for times 6-8am, 8-10am, 10-12noon, 12-2pm and 6pm-6am respectively. Error bars are all set at the 85% confidence level and points are displayed at the midpoint of each time interval.

Not surprisingly, these results show that eye localization plays a significant role in the accuracy of face recognition. Note that excluding images in which faces were not found at all by FaceIt did not significantly affect the measured difference.

One point to note, instances in which there is a difference in FaceIt performance based on computed eye locations indicates a difference in higher stages of FaceIt processing, since the only difference in the face recognition algorithm processing is the accuracy of the eye localization.



(a)



(b)

Figure 9: Performance comparison for FaceIt using eye coordinates obtained by FaceIt vs. eye coor-dinates computed from markers on "weathered" face images for (a) NEAR camera and (b) FAR camera.

Incidentally, the rather large difference in performance for evening images was due to the larger dynamic range of the far camera images as compared to the near camera images, likely enabling more effective face image processing.

5. ATMOSPHERIC IMAGE DEGRADATION

The following section describes an experiment performed to investigate the effect of eye localization on images with subjects under various illumination conditions and poses.

5.1 Experiment Conditions and Factors

A total of 68 individuals from the CMU PIE database with neutral expressions, under 21 illuminations and 3 poses were used in this experiment [14]. Using front profile images in the gallery with room lights on, FaceIt performance was measured on probes each consisting of the 68 individuals under 21x3=63 different combinations of illumination and pose. Illuminations resulting in the darkest images were not used to insure that FaceIt would be able to properly locate the face.

Poses used were front profile and the two closest poses on either side of the front profile. FaceIt was run twice, once using FaceIt eye localization, and a second time using specified eye coordinates.

5.2 Experiment Conditions and Factors

As before, the cumulative match score (CMS) for rank 1 was used as a measure of performance. However, in this case, balanced repeated replication (BRR) was used to tighten confidence intervals over the small set of data [8].

5.3 Results and Discussion

Results are shown in figure 10 for three poses, with CMS values averaged over all illuminations using the BRR technique. As shown in the plot, there is a significant difference in performance for poses different from the front profile view (at a 90% confidence level). Poor eye localization clearly degrades the performance of the algorithm on the probe images.

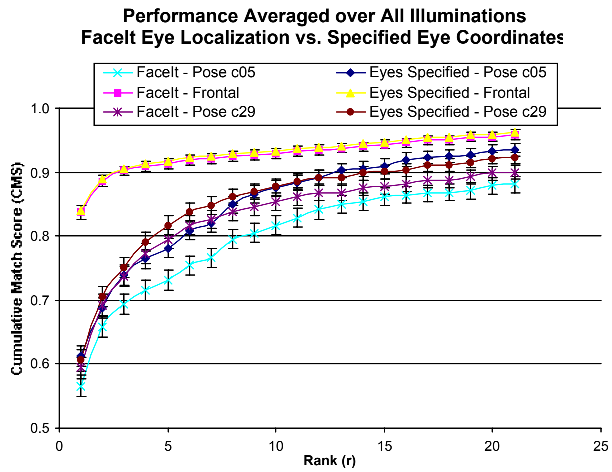


Figure 10: Traditional FERET CMS curve for three poses, one set using FaceIt eye localization, the other using computed eye locations.

6. CONCLUSION

While many people would expect eye localization to have an impact on recognition accuracy, this paper has shown, for the first time, that it has a significant quantitative impact. Even with ideal data, it was shown to have a significant effect on the recognition accuracy of several different face recognition algorithms. The response of the various face recognition algorithms to eye perturbations was found to be similar, despite significant differences in algorithm design, suggesting that our observations are relevant to many face analysis systems.

The simulation using ideal data was then validated using the same experimental technique with actual images. The behavior of the algorithms did not alter drastically when the same eye perturbations were used on "real" images, although the degraded images did result in performance loss being somewhat attenuated. This was, at first, somewhat surprising since one might expect eye localization to increase in importance as the quality of the face image degrades. Our results indicate that blurring/noise from the real imaging system may in fact *help* to make the algorithm less sensitive to errors in eye location. The working hypothesis is that blurring seems to mitigate the effect of eye localization, making (at least) these algorithms more stable with respect to eye perturbations. This may explain the observation that sometimes blurring helped recognition [5] and is consistent with human experiments that show that face processing is less affected by blurring than expected [6]; blurring may impact eye localization, but image smoothing mitigates the impact of such errors.

Results comparing the performance of FaceIt using its own eye-localization algorithm versus known eye coordinates on real face image data (both atmospherically degraded and under various pose and illumination conditions) showed that even a leading recognition algorithm can be significantly improved by increasing the accuracy of its eye localization routines.

The importance of eye *separation* versus eye *location* was also noted. Our observations suggest that reliance on eye separation for scaling and normalization may be precarious especially in instances where face images are degraded or taken in more realistic environments. Since much of later face processing relies on proper scaling, our results suggests the need for additional "backup" methods of scaling, which may either complement the use of eye separation or supercede it.

7. ACKNOWLEDGMENTS

Our authors would like to thank Derek Drake, Walter Scheirer, and Shanna Terry for their diligence and perseverence in generating and processing the data used in these experiments. This work was supported in part by the DARPA HID program, ONR contract # N00014-00-1-0388. The views herein are those of the authors and do not necessarily represent those of the funding agencies.

8. REFERENCES

[1] Batur, Aziz U. and Flinchbaugh, Bruce E. (2001). Performance Analysis of Face Recognition Algorithms on TMS320C64x. *Application Report*, DSP Solutions R&D Center.

[2] Bolme, D.S., Beveridge, J.R., Teixeira, M. and Draper, B.A. (2003). The CSU Face Identification Evaluation System: Its Purpose, Features, and Structure. *ICVS 2003*: 304-313.

[3] Brunelli, R. and Poggio, T. (1993). Face Recognition: Features vs. Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042-1052.

[4] Chellappa, R., Wilson, C.L., and Sirohey, S. (1995). Human and Machine Recognition of Faces: A Survey. *Proc. of the IEEE*, 83(5):705-740.

[5] Heseltine, T.,Pears, N. and Austin, J. (2002). Evaluation of Image Pre-processing Techniques for Eigenface Based Face Recognition. Tech. Report, Dept. of Comp. Science, Univ. of York.

[6] Hole, G.J., George, P.A., Eaves, K. and Rasek, (2002). A. Effects of Geometric Distortions on Face-recognition Performance. *Perception* 31(10):1221-1240.

[7] Huang, C.L. and Chen, C.W. (1992). Human facial feature extraction for face interpretation and recognition. *Pattern Recognition,* 25(12):1435-1444.

[8] Micheals, R. J. and Boult, T.E. (Dec. 2001) Efficient Evaluation of Classification and Recognition Systems*. Proc. of the IEEE Conf. on Comp. Vision and Patt. Rec.* pp. 50-57.

[9] Mu, F., Li, H.Y., and Forchheimer, R. (1996). Automatic Extraction of Human Facial Features. *Signal Processing: Image Comm.* 8(4):309-326.

[10] Okada, K., Steffens, J., Maurer, T., Hong, H., Neven, H. and von der Malsburg, C. (1998). The Bochum/USC Face Recognition System and How It Fared in the FERET Phase III Test. In *Wechsler et al.*, eidtores*, Face Recognition: From Theory to Applications*, pp. 186-205.

[11] Penev ,P. S. and Atick, J. J. (1996). Local feature analysis: A general statistical theory for object representation. *Neural Systems*, 7:477-500.

[12] Phillips, P.J., Grother, P., Micheals, R.J., Blackburn, D.M., Tabassi, E., and Bone, M. (2003). Face Recognition Vendor Test 2002 (FRVT 2002). Technical Report NISTIR 6965, NIST.

[13] Phillips, P.J., Wechsler, H., Huang, J.S., and Rauss, P.J. (1998). The FERET Database and Evaluation Procedure for Face-Recognition Algorithms. *Image & Vision Comp.* 16(5):295-306.

[14] Sim, T., Baker, S. and Bsat, M. (2002). The CMU Pose, Illumination and Expression (PIE) Database. In *Proc. of the IEEE Int. Conference on Automatic Face and Gesture Recognition*.

[15] Stringa, L. (1993). Eyes detection for face recognition. *Appl. Art. Intelligence*, 7(4):365-382.

[16] Takács, B. and Wechsler, H. (1997). Detection of Faces and Facial Landmarks Using Iconic Filter Banks. *Pattern Recognition*, 30(10):1623-1636.

[17] Turk, M.A. and Pentland A.P. (1991). Face Recognition Using Eigenfaces. In *Proc. of IEEE Conf. on Comp. Vision and Patt. Rec*., pp. 586-591.